



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H

Citation for published version:

Makou, E, Mertens, HDT, Maciejewski, M, Soares, DC, Matis, I, Schmidt, CQ, Herbert, AP, Svergun, DI & Barlow, PN 2012, 'Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H', *Journal of Molecular Biology*, vol. 424, no. 5, pp. 295-312.
<https://doi.org/10.1016/j.jmb.2012.09.013>

Digital Object Identifier (DOI):

[10.1016/j.jmb.2012.09.013](https://doi.org/10.1016/j.jmb.2012.09.013)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Molecular Biology

Publisher Rights Statement:

Open Access funded by Wellcome Trust

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Solution Structure of CCP Modules 10–12 Illuminates Functional Architecture of the Complement Regulator, Factor H

Elisavet Makou¹, Haydyn D. T. Mertens², Mateusz Maciejewski¹, Dinesh C. Soares³, Ilias Matis¹, Christoph Q. Schmidt¹, Andrew P. Herbert¹, Dmitri I. Svergun² and Paul N. Barlow¹

1 - Schools of Chemistry and Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK

2 - European Molecular Biology Laboratory Hamburg Outstation, c/o Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22603 Hamburg, Germany

3 - Medical Genetics Section, Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK

Correspondence to Paul N. Barlow: Paul.Barlow@ed.ac.uk

<http://dx.doi.org/10.1016/j.jmb.2012.09.013>

Edited by C. Kalodimos

Abstract

The 155-kDa plasma glycoprotein factor H (FH), which consists of 20 complement control protein (CCP) modules, protects self-tissue but not foreign organisms from damage by the complement cascade. Protection is achieved by selective engagement of FH, via CCPs 1–4, CCPs 6–8 and CCPs 19–20, with polyanion-rich host surfaces that bear covalently attached, activation-specific, fragments of complement component C3. The role of intervening CCPs 9–18 in this process is obscured by lack of structural knowledge. We have concatenated new high-resolution solution structures of overlapping recombinant CCP pairs, 10–11 and 11–12, to form a three-dimensional structure of CCPs 10–12 and validated it by small-angle X-ray scattering of the recombinant triple-module fragment. Superimposing CCP 12 of this 10–12 structure with CCP 12 from the previously solved CCP 12–13 structure yielded an S-shaped structure for CCPs 10–13 in which modules are tilted by 80–110° with respect to immediate neighbors, but the bend between CCPs 10 and 11 is counter to the arc traced by CCPs 11–13. Including this four-CCP structure in interpretation of scattering data for the longer recombinant segments, CCPs 10–15 and 8–15, implied flexible attachment of CCPs 8 and 9 to CCP 10 but compact and intimate arrangements of CCP 14 with CCPs 12, 13 and 15. Taken together with difficulties in recombinant production of module pairs 13–14 and 14–15, the aberrant structure of CCP 13 and the variability of 13–14 linker sequences among orthologues, a structural dependency of CCP 14 on its neighbors is suggested; this has implications for the FH mechanism.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

The activated complement system^{1–3} assists in clearance of pathogens in addition to debris from diseased, damaged, dead or dying cells. Hydrolysis of a thioester linkage in complement component C3, yielding C3(H₂O), is the rare, spontaneous but ubiquitous initiating event of the “alternative” pathway of the complement cascade.⁴ C3(H₂O) collaborates with proenzyme factor B and proteolytic factor

D to create the convertase complex C3(H₂O).Bb that converts C3 to C3a (an anaphylatoxin) and C3b. C3b is able to rapidly attach covalently to any nearby surface and to form C3b.Bb [a structural and functional analogue of C3(H₂O).Bb] that, in a positive-feedback loop,⁵ converts additional C3 to C3b. Particles become coated (“opsonised”) with C3b and are then subject to immune clearance. Moreover, C3b participates in both “classical” and “lectin” pathways; all three pathways feed into the

terminal pathway that releases the anaphylatoxin C5a and culminates in the assembly of potentially cytolytic pores.

Distinction by the alternative pathway between self and nonself is not driven by discriminatory covalent attachment of C3b.^{4,6} Rather, it is accomplished mainly by selective action of factor H (FH), complement receptor type 1 (CR1), membrane cofactor protein (MCP) and decay accelerating factor (DAF).⁷ Either through being host-membrane bound (CR1, DAF or MCP) or (in the case of FH) via selective association with host-cell-specific markers,⁸ these regulators prevent amplification of C3b on self-surfaces. They (DAF, CR1 and FH) perturb formation and stability of C3b.Bb or (CR1, FH and MCP) recruit factor I to cleave C3b, removing it from the amplification loop that continues to operate on, for example, a bacterium. The product, iC3b, cannot form a complex with Bb but remains an opsonin and a ligand for receptors on immune cells.⁹

The abundant¹⁰ plasma regulator FH (155 kDa, 1213 residues)^{11,12} fails to adequately protect host tissues from complement-mediated damage over the lifetimes of many individuals who possess *CFH* polymorphisms or mutations.¹³ Age-related macular degeneration is a prominent example of a disease arising from such a deficiency.¹⁴ To explain the molecular bases for these disorders, it is essential to understand why wild-type FH is significantly more active on healthy self-cells than on diseased or foreign ones.¹⁵ Moreover, as a self-tissue-specific regulator of C3 activation, knowledge of the structure–function relationships of FH¹⁶ could inspire design of therapeutic complement inhibitors useful in the wide range of conditions where complement plays a damaging role.^{13,17–19}

FH consists entirely of 20 complement control protein (CCP) modules²⁰ of between 51 and 62 residues each, connected by linking sequences of between 3 and 8 residues in length^{21,22} (Fig. 1). Successful crystallization of FH has not been reported. Current structural knowledge draws on data from small-angle X-ray scattering (SAXS), analytical ultracentrifugation and electron microscopy^{23–26} combined with high-resolution structures for short discontinuous segments of the protein.^{27–36}

The N-terminal CCPs 1–4 and the C-terminal CCPs 19–20 of FH bind C3b at adjacent, nonoverlapping sites.^{34,37,38} Like full-length FH, a recombinant fragment consisting solely of CCPs 1–4 (i.e., FH 1–4) not only competes with binding of factor B to C3b in solution but also recruits factor I.^{39–42} In addition, FH has sites within CCPs 6–8 and CCPs 19–20 that interact with polyanionic self-surface markers and are critical for specificity.^{37,43–46} Nearly all disease-linked sequence variations occur at these C3b-binding or polyanion-binding regions of FH¹³ in which CCPs adopt extended end-to-end arrangements connected by linkers of three or four

residues. The central modules of FH (CCPs 9–18), on the other hand, do not contain strong binding sites for polyanions or C3b.³⁷ Thus, tethering together the two C3b/polyanion-binding sites, CCPs 1–8 and CCPs 19–20, might produce an efficient biopharmaceutical.⁴⁷ However, knowledge on the role of the intervening 10 modules of FH that are hypothesized to enhance the specificity of FH for self-surface protection is lacking.

This central region has long inter-modular linkers (6, 6, 8, 7 and 5 residues for CCPs 10–11, 11–12, 12–13, 13–14 and 14–15, respectively) (Fig. 1), relatively small modules (51 residues in CCP 13), and accommodates seven of eight utilized N-glycosylation sites in FH.⁴⁸ Initial notions that CCPs 9–18 merely form a flexible tether between C3b-binding and polyanion-recognition regions were disproved by studies of FH 10–15.³⁵ This molecule failed to crystallize, but a SAXS-derived structural model guided by an NMR-derived FH 12–13 structure was created. The FH 10–15 construct forms a compact rather than flexible or elongated structure.

Here, we describe recombinant production of the remaining bi-modules (i.e., besides FH 12–13) from the FH 10–15 region and their structural investigation. While samples of FH 13–14 and FH 14–15, unusually for CCP bi-modules, resisted all attempts at preparation in our hands, structures of FH 10–11 and FH 11–12 were readily determined by NMR. A structure of FH 10–12, produced by combining these two high-resolution solution structures, was validated by SAXS. A subsequently constructed model of FH 10–13 allowed reinterpretation of SAXS data for FH 10–15 and FH 8–15. It seems that CCPs 13–15 are compacted, while CCP 14 appears unstable unless it is folded back along CCPs 12 and 13. These possibilities are discussed in terms of the mechanism of FH.

Results

FH 10–11 and FH 11–12, but not FH 13–14 or FH 14–15, yielded assignable NMR spectra

We set out to investigate whether the large tilt angle (~80°) previously observed between CCPs 12 and 13³⁵ is a consistent feature across the central segment of FH. Like FH 12–13, samples of FH 10–11 and FH 11–12 were straightforward to prepare. Moreover, they gave [¹H, ¹⁵N]heteronuclear single quantum coherence (HSQC) spectra of high quality with the expected numbers of well-dispersed cross-peaks (Fig. 2a and b) for stably folded proteins.

Other overlapping bi-modules within the FH 10–15 region—FH 13–14 and FH 14–15—proved intractable. Recombinant FH 13–14 was prone to proteolysis and self-association; nevertheless, a ¹⁵N-

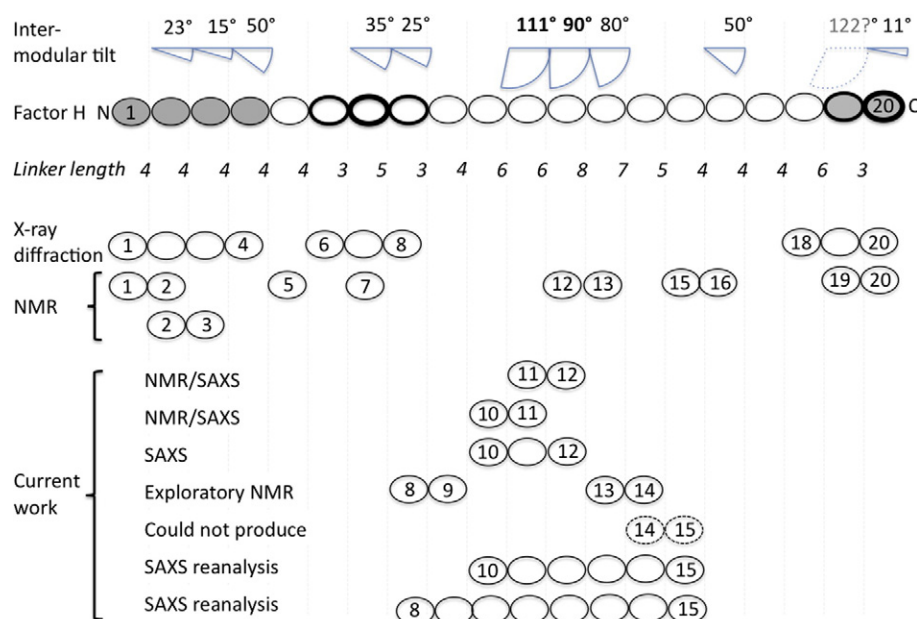


Fig. 1. Gaps in structural knowledge of human complement FH. Drawn as ovals, CCPs are shaded to indicate C3b-binding regions or emboldened to signify GAG/sialic acid-recognition sites (CCPs 7 and 20 being the most strongly implicated). The number of residues in each inter-modular linker is written beneath; above is a schematic representation of the tilt angle (away from linearity) of a CCP's long axis with respect to the long axis of the preceding CCP [including (bold) FH 10–11 and FH 11–12 tilts measured herein]. The tilt between CCPs 18 and 19 was 122° in the FH 18–20 crystal structure,³⁶ but SAXS suggested a more extended arrangement in solution. Structures solved by NMR or X-ray crystallography of recombinant segments of FH are summarized in the central portion of the figure. “Current work” summarizes the recombinant constructs investigated and methods used for analysis.

labeled sample eluting as a monomeric species was obtained from a calibrated size-exclusion chromatography column and produced a single band, of the expected size, by sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Unexpectedly, this FH 13–14 sample must have been improperly folded since, under a range of conditions, it yielded [¹H,¹⁵N]HSQC spectra (Supplementary Fig. S1) dominated by overlapped, broad cross-peaks. An overlay of this spectrum with a [¹H,¹⁵N]HSQC spectrum of FH 13 (previously expressed as an isolated module³⁵) showed that only a few weak cross-peaks from compactly folded CCP 13 could be detected in the CCP 13–14 spectrum (Supplementary Fig. S1). In the case of FH 14–15, no sample suitable for NMR data collection could be prepared due to excessive propensity for proteolysis and oligomerization.

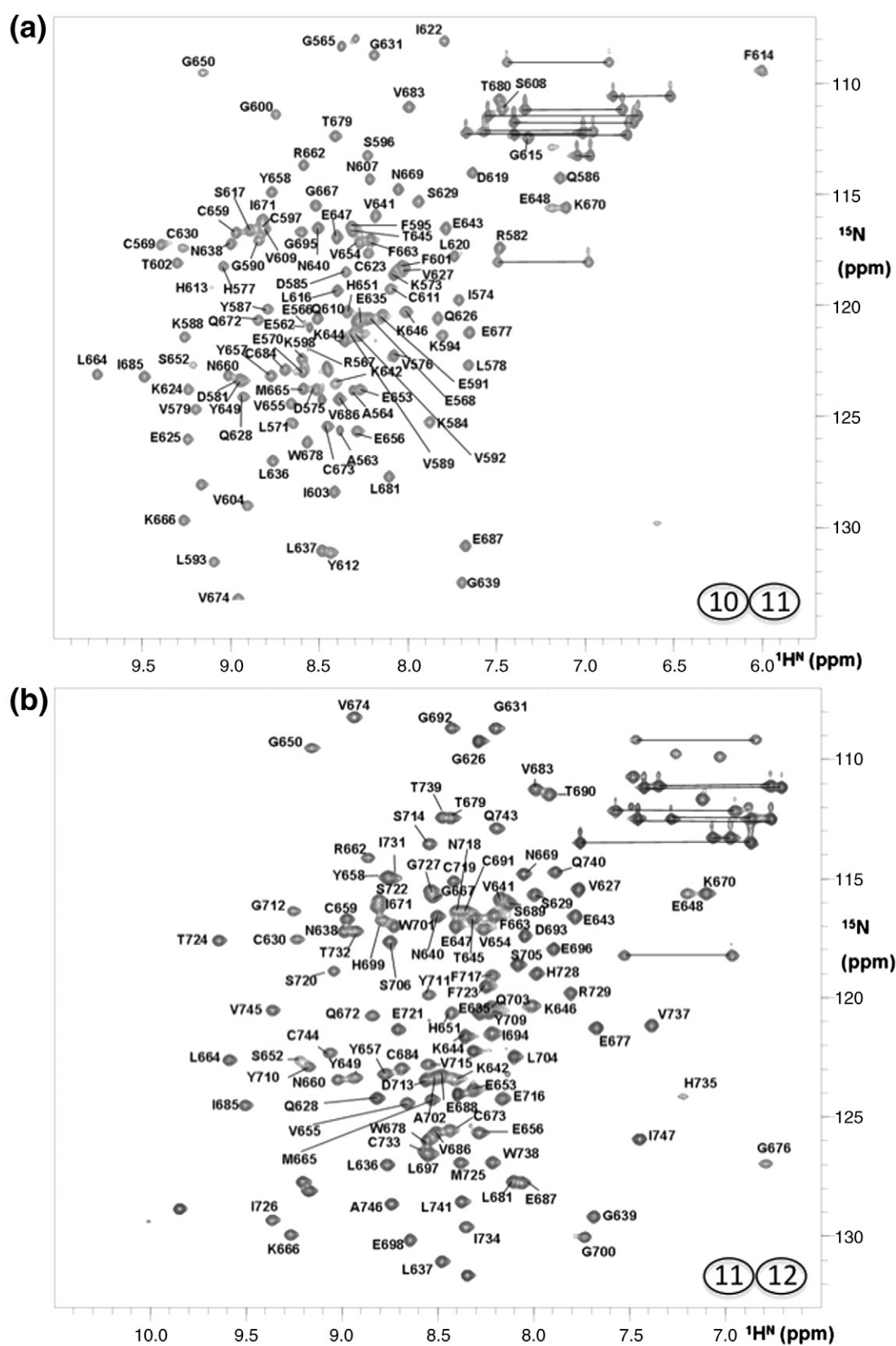
Modules 10–12 do not form a highly elongated structure

SAXS curves were collected for FH 10–11, FH 11–12 and FH 10–12; neither FH 13–14 nor FH 14–15 samples were suitable. Analysis of parameters extracted from scattering profiles and real-space distance distribution functions, $p(r)$, reveals mono-

mers (Fig. 3 and Table 1) consistent with their elution volumes from size-exclusion chromatography columns. Bi-modules FH 10–11 and FH 11–12 yielded similar radii of gyration (R_g) and maximum particle dimensions (D_{max}), but FH 11–12 ($R_g=2.4$ nm and $D_{max}=8.3$ nm) was slightly more extended than FH 10–11 ($R_g=2.2$ nm and $D_{max}=7.5$ nm). The positively skewed $p(r)$ functions of both bi-modules are characteristic of dumbbell-shaped structures; the first maximum represents a common average intra-module distance (~1.7 nm), and the second maximum represents an approximate average separation of the centers of mass of the two CCPs (~3.8 nm). Studies of tri-modular FH 10–12 yielded only a modest 0.1- to 0.3-nm increase in R_g and 0.8- to 1.2-nm increase in D_{max} (Fig. 3 and Table 1). The $p(r)$ function of FH 10–12 shows a shift in the position of the first maximum relative to that of bi-modules; this reflects an increase in effective cross-section upon addition of the third CCP. These observations clearly demonstrate that FH 10–12 does not have an elongated structure similar to FH 1–3³³ or FH 6–8.³²

Three-dimensional structures of FH 10–11 and FH 11–12

The NMR spectra (¹³C, ¹⁵N) for FH 10–11 and 11–12 (Fig. 2a and b) enabled near-complete assignment



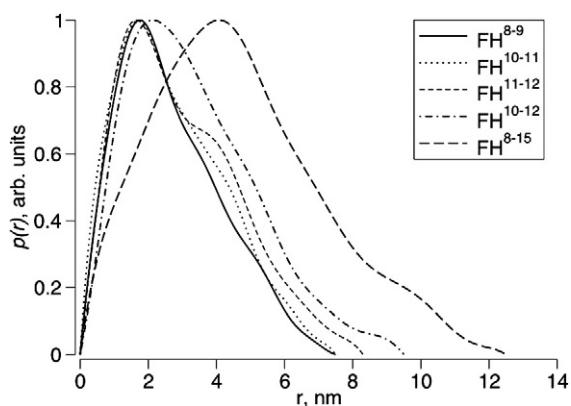


Fig. 3. SAXS data for FH constructs: pairwise interatomic distance distributions, $p(r)$, for the following constructs derived from FH, FH 8–9, FH 10–11, FH 11–12, FH 10–12 and FH 8–15.

of ^1H , ^{15}N and ^{13}C (Table 2). Following manual partial assignment of ^{15}N - and ^{13}C -filtered nuclear Overhauser enhancement (NOE) spectroscopy spectra, we input a provisional list of NOE assignments into the first of seven CYANA cycles of structure calculations.⁴⁹ CYANA-derived distance restraints were used for calculations of structures in CNS (Crystallography and NMR System).⁵⁰ The 20 lowest-energy CNS-derived structures converged well and were refined in explicit aqueous solvent (Fig. 4a and b; Table 2). Note (as discussed further below) that near-identical structures resulted from a calculation that incorporated both NOEs and SAXS-derived parameters.

The β -strand-rich structures of CCPs 10, 11 and 12 (Fig. 4c and d) contain Cys(I)–Cys(III) and Cys(II)–Cys(IV) disulfides. The consensus Trp is buried in the cores of CCPs 11 and 12; a Leu occupies this position in CCP 10. STRIDE-identified⁵¹ β -strands were labeled (in Fig. 4c and d) according to a convention recognizing the occurrence of maximally eight β -strands (A–H), at conserved positions, within CCP structures.^{29,52} Slowly exchanging amide protons (Supplementary Fig. S2) coincided with either burial or engagement in hydrogen bonds; after 15 min of D_2O exposure, the following numbers of HSQC cross-peaks remained: 2 in CCP 10 and 22 in CCP 11 (none in the linker) for FH 10–11, and 17 in CCP 11 and 9 in CCP 12 (1 in the linker) for FH 11–12. Although CCP 10 has more extensive β -strands, CCPs 11 and 12 have more slowly exchanging amides.

Structures of CCP 10 or 11 converged well when the ensemble of NOE-derived FH 10–11 structures was overlaid solely on either module; CCP 11 or 12

within the FH 11–12 ensemble likewise converged well on a per-module basis (Fig. 4a and b). The structure of CCP 11 from FH 10–11 overlaid well with the structure of CCP 11 from the FH 11–12 structure; moreover, CCP 12 solved in the current work overlaid well on CCP 12 from FH 12–13 (Supplementary Fig. S3 and Table S1). That module structures were conserved in FH 10–11, FH 11–12 and FH 12–13 endorsed construction of models of longer FH segments by concatenating bi-module structures (see below). While CCP 10's structure does not resemble the structures of other FH CCPs (backbone rmsd $> 2 \text{ \AA}$; Table S1), it is similar to MCP CCP 3⁵³ (Supplementary Fig. S4a). CCP 12 is very similar in structure to CCP 2 of complement receptor type 2⁵⁴ (Supplementary Fig. S4b). FH CCPs 11 and 12 resemble one another and are similar to FH CCPs 18 and 19³⁶ (Supplementary Fig. S4c and Table S1).

NOE-derived structures are not under-restrained or over-restrained

Within each bi-module, ~ 120 NOEs were identified between modules and their linker (Table 2). Eight NOEs were found directly between CCPs 10 and 11, compared to 38 NOEs between CCPs 11 and 12; thus, a network of NOEs that was more extensive for FH 11–12 than FH 10–11 defines the inferred inter-module orientations. Indeed, the FH 10–11 structures exhibited good convergence when overlaid over both modules, and a comparable result was obtained for FH 11–12. Moreover, relaxation data (Supplementary Fig. S5) showed that inter-module linkers are not more mobile than other extended regions of the polypeptide. Nonetheless, several further steps were taken to check that these FH 10–11 and FH 11–12 solution structures were neither over-restrained nor under-restrained.

First, bi-module structures were recalculated using an ensemble-based simulated annealing protocol (in Xplor-NIH).^{55,56} This allows simultaneous generation of an ensemble of n conformers that, between them, satisfy all NOE-derived distance restraints. With the use of an ensemble size (n) of 2, ensemble members turned out to be nearly identical with one another and to the outcome of a standard simulated annealing protocol (i.e., effectively, $n=1$); thus, we found no evidence that FH 10–11 or FH 11–12 adopts multiple conformers interconverting on the NOE timescale.

Second, NOE (only)-derived structures of FH 10–11 were fit to the FH 10–11 SAXS data (using CRY SOL⁵⁷) with discrepancies (χ) (between data

Fig. 2. NMR data ($[\text{}^1\text{H}, \text{}^{15}\text{N}]$ HSQC spectra) for FH 10–11 and FH 11–12. Spectra [(a) FH 10–11; (b) FH 11–12] were collected in 30 min on double-labeled samples (0.45 mM and 0.65 mM FH 10–11 and FH 11–12, respectively) at 25°C in 20 mM potassium phosphate buffer (at pH 6.7 and 6.3 for FH 10–11 and FH 11–12, respectively).

Table 1. SAXS parameters for FH fragments

Construct	R_g^{autoRg} (nm)	R_g^{GNOM} (nm)	D_{max} (nm)	V_p (nm ³)	MM _{SAXS} (kDa)	V_{DAM} (nm ³)	MM _{DAM} (kDa)
FH 8–9 (13.7 kDa)	2.1±0.1	2.2±0.1	7.5±0.5	21±5	15±5	23±5	11±5
FH 10–11 (13.9 kDa)	2.2±0.1	2.2±0.1	7.5±0.5	22±5	16±5	21±5	10±5
FH 11–12 (14.1 kDa)	2.4±0.1	2.3±0.1	8.3±0.5	25±5	19±5	25±5	12±5
FH 10–12 (20.6 kDa)	2.7±0.1	2.7±0.1	9.5±0.5	33±5	20±5	47±5	24±5
FH 8–15 (54.7 kDa)	3.5±0.1	3.7±0.1	12.5±0.5	92±10	62±5	93±10	47±5

R_g^{autoRg} and R_g^{GNOM} are the radii of gyration estimated from the SAXS data using the automated Guinier analysis routine and GNOM, respectively. D_{max} , V_p , V_{DAM} , MM_{SAXS} and MM_{DAM} are the maximum particle dimension, hydrated particle volume, dummy-atom model total excluded volume, molecular mass estimated from $I(0)$ and molecular mass estimated from the dummy-atom model volume, respectively. The data shown are averaged or merged and extrapolated to infinite dilution, using the SAXS profiles recorded at several concentrations. The molecular mass calculated from the sequence is shown in parentheses.

and model) of 1.3–1.6; equivalent values for FH 11–12 were 1.6–2.2 (Supplementary Fig. S6a). Thus, for both bi-modules, the NMR-derived structures provide a good fit to the scattering data. Furthermore, in an *ab initio* approach, the average structures of both bi-modules were reconstructed from SAXS data (using DAMMIF⁵⁸). Within each set of 10 independent reconstructions, highly similar shapes were obtained, with low (<1.0) normalized spatial discrepancies.⁵⁹ Importantly, discrepancies between these models and high-resolution NMR-derived structures were small ($\chi=0.9$ –1.2) in both cases. When averaged and volume-fitted *ab initio*

shapes were superimposed with the corresponding NMR ensembles using SUPCOMB,⁵⁹ good spatial agreement was observed (Supplementary Fig. S6b), thus validating the NMR-derived structures.

Finally, structure calculations (with ensemble sizes of 1 and 2) were conducted for each bi-module with incorporation of both NOE and SAXS data⁶⁰ as experimental structural restraints.⁶¹ These calculations yielded structures (Table S2) that were nearly identical with NOE-only structures and exhibited no NOE violations >0.5 Å.

CCPs 11 and 12 are more intimately associated than CCPs 10 and 11

Bi-modules FH 10–11 and FH 11–12 are bent (or tilted) by ~118° and ~90°, respectively, away from a linearly extended form. The surface area effectively buried between the N-terminal (CCP 10) and the C-terminal (CCP 11) halves of FH 10–11 (using the peptide bond between the middle two residues of the linker as a boundary) is only ~360 Å², consistent with detection of just eight NOEs between these modules. An inferred salt bridge (Supplementary Fig. S7) links the first residue of the linker (Lys624, the last residue of CCP 10 β-strand H) to Asp675 of the CCP 11 FG β-turn. Residues Thr602 and Val604 from the CCP 10 β-strand E (paired with CCP 10 β-strand H) participate in inter-modular van der Waals contacts with CCP 11 residues His651 and Asp675 (in BD loop and FG β-turn, respectively). All four contribute, along with alkyl portions of linker residues Lys624, Val627 and Gln628, to a hydrophobic pocket wedged between the modules (Fig. 5a). Polar side chains of linker residues Glu625, Gln626 and Ser629 are solvent exposed on the convex surface of the inter-modular bend.

In FH 11–12, side chains of the DE loop and β-strand E of CCP 11 (Pro661, Arg662, Phe663 and Leu664) make extensive van der Waals interactions with alkyl portions of linker residues (while polar portions are displayed on the convex surface of the bend) and with a segment of the BD loop of CCP 12 (Tyr710, Tyr711 and Gly712) (Fig. 5b). All these side chains are substantially buried in a hydrophobic

Table 2. Statistics for solution structures of FH 10–11 and FH 11–12

Segment	FH 10–11	FH 11–12
Numbers of NOE-derived distance restraints		
Intraresidue, $ i-j \leq 1$	1405	1482
Medium range, $1 < i-j < 5$	411	301
Long range, $ i-j \geq 5$	1391	1322
Total	3207	3105
Inter-modular ^a	8	38
N-module to linker	62	89
C-module to linker	65	39
rmsd values (Å) (superimposition of ensemble over entire sequence)		
All atoms	1.17	1.05
Backbone atoms	0.91	0.84
rmsd values (Å) (superimposition of ensemble over N-terminal CCP)		
All atoms	0.79	0.73
Backbone atoms	0.50	0.33
rmsd values (Å) (superimposition of ensemble over C-terminal CCP)		
All atoms	0.74	0.72
Backbone atoms	0.33	0.43
Inter-modular angles (°) [minimum–maximum (mean±SD)]		
Skew	75–107 (89±8)	149–170 (158±4)
Twist	40–78 (57±10)	59–87 (72±7)
Tilt	111–128 (118±5)	76–99 (88±6)
Ramachandran assessment (%)		
Most favored	80.5	74.6
Additionally allowed	16.4	22.4
Generously allowed	1.3	1.7
Disallowed	1.8	1.3
Surface area buried between modules (Å ²) ^b	358	642

^a Module boundaries defined by the first and the last cysteine of each CCP module.

^b This value is for the closest-to-mean structures in each ensembles (see Materials and Methods).

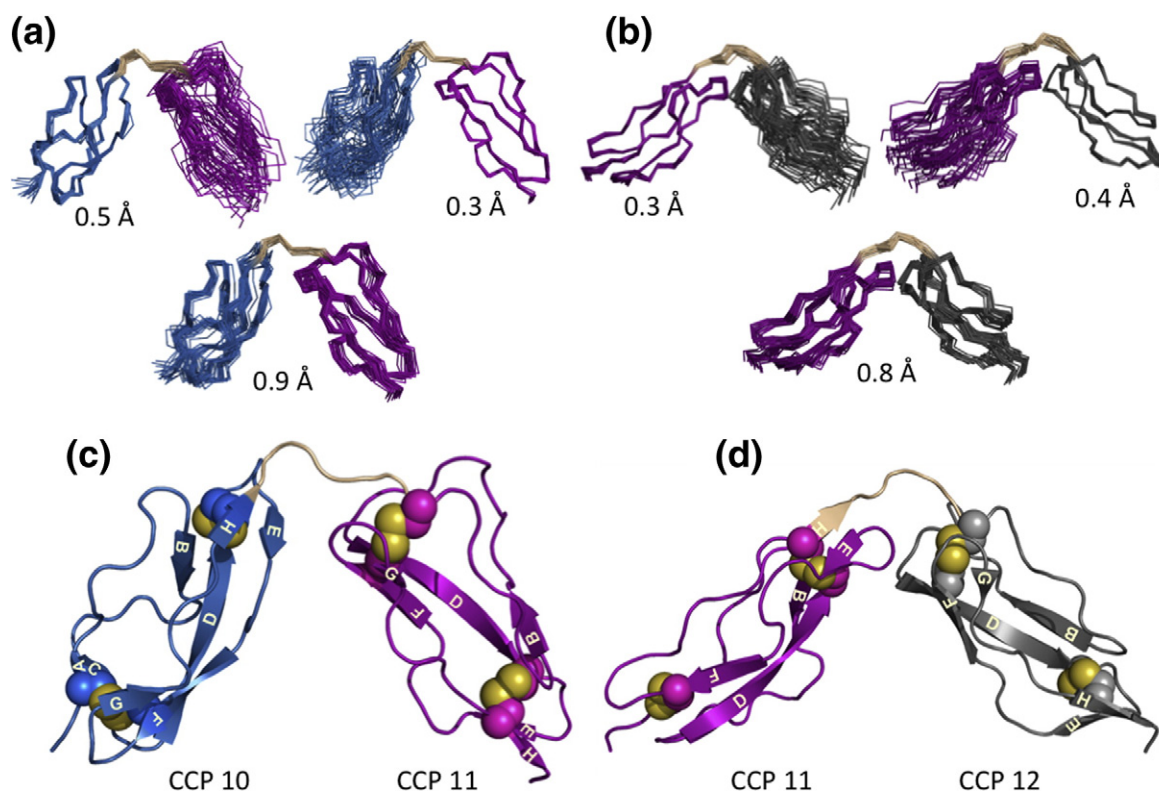


Fig. 4. Solution (NOE-derived) structures of FH 10–11 and FH 11–12. (a) The 20 lowest-energy, water-minimized FH 10–11 structures, overlaid on backbone atoms of CCP 10 (blue), CCP 11 (magenta) or both modules CCP 10 and 11 (and linker; wheat) (relevant rmsd values are shown). (b) As (a) except that NOE-derived structures of FH 11–12 (CCP 12 in gray) are overlaid. In the cartoon representations of (c) FH 10–11 and (d) FH 11–12, β -strands identified by STRIDE were labeled A–H according to a convention whereby CCP modules have a maximum of eight β -strands. Spheres represent cysteine side chains (sulfurs are colored yellow).

pocket significantly bigger than the equivalent pocket in FH 10–11—the surface area buried between modules 11 and 12 is $\sim 640 \text{ \AA}^2$.

In summary, CCPs 11 and 12 are welded together at an $\sim 90^\circ$ tilt in the structure calculation by a substantial number of mutually compatible NOEs reflecting extensive interactions and burial of a sizeable surface area. Modules 10 and 11 are mutually orientated at an even greater bend angle to accommodate (NOE-supported) interactions between strand E of CCP 10 and the CD and FG loops of CCP 11. Only a small surface area is buried between CCPs 10 and 11, but a Lys624–Asp675 salt bridge may stabilize it.

FH 10–12 can be modeled from SAXS data and concatenation of bi-modules

Samples of FH 10–12 did not yield NMR spectra suitable for structure determination. This problem, encountered in previous efforts to solve triple-CCP-module structures,^{33,52} was ascribed to anisotropic tumbling yielding lower T_2 values than a globular protein of comparable mass.

Broad peaks and crowded spectra precluded identification of sufficient inter-modular and module-to-linker NOEs without more elaborate isotopic labeling. Nonetheless, FH 10–12 yielded useful SAXS data. Three methods were used to combine results of SAXS for FH 10–12 and NMR for FH 10–11 and FH 11–12.

First, a model of FH 10–12 built by concatenating FH 10–11 and FH 11–12 structures (Fig. 6a and b) was validated on the basis of successful fitting to SAXS data for FH 10–12 ($\chi=0.8$) (Fig. 6c) using the same method (see above) as employed for bi-modules. Second, in an *ab initio* approach, also similar to that described for the bi-modules, the average solution structure of FH 10–12 was reconstructed from SAXS data. As with the bi-modules, independent reconstructions yielded similar shapes (average normalized spatial discrepancies = 0.73 ± 0.12), while discrepancies between models and experimental data were low ($\chi = 0.88 \pm 0.01$), whereas the total excluded dummy-atom volume of the model matched the expected monomeric molecular mass. Encouragingly, the average *ab initio* shape superimposed well with, and thereby validated, the

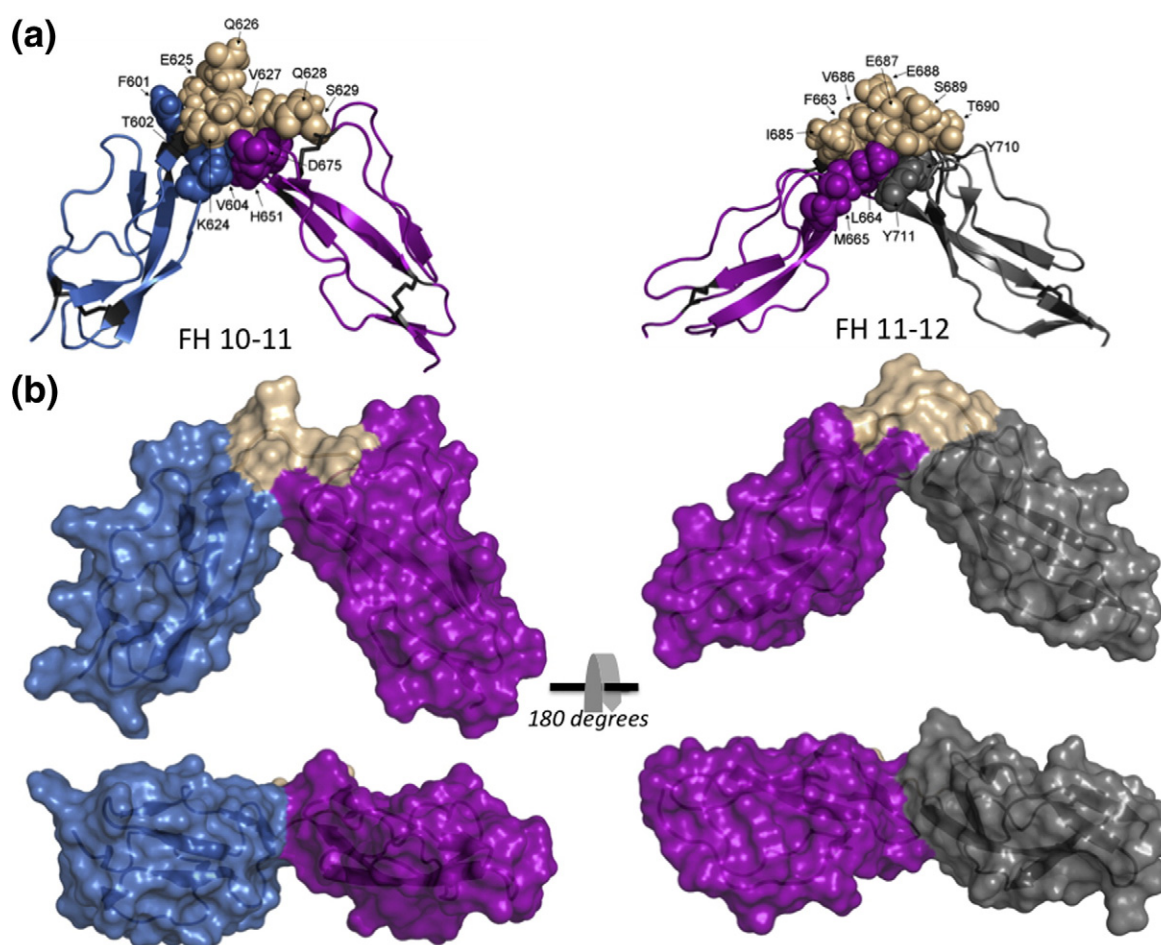


Fig. 5. Inter-module interfaces in FH 10–11 and FH 11–12. (a) Side chains of residues involved in the interfaces drawn as color-coded labeled spheres: blue for CCP 10, magenta for CCP 11 and gray for CCP 12; wheat for the CCP 10–11 or the CCP 11–12 linker. (b) With the use of identical color coding, solvent-accessible surfaces are drawn, illustrating more extensive contacts between CCPs 11 and 12 despite larger tilt angle between CCPs 10 and 11.

concatenated NMR model of FH 10–12 (Supplementary Fig. S2b). Finally, a combined FH 10–12 NOE list was compiled by concatenating FH 10–11 and FH 11–12 NOEs in a way that eliminated duplication or conflict of NOEs originating from the mutual CCP 11. Then, FH 10–12 structures were calculated using simulated annealing, employing as restraints both SAXS data and the combined NOE list. As with bi-modules, ensembles ($n=1$ or 2) of structures were derived, with energy terms calculated as ensemble averages. Calculated FH 10–12 structures were in good agreement with the concatenated model, with each other and with both NOE and SAXS data.

The ensemble optimization method (EOM)⁶² was used to investigate flexibility of FH 10–12. In an initial application, NMR-derived structures of individual CCPs 10, 11 and 12 were used as rigid bodies and linker regions defined as flexible chains of dummy atoms to generate a pool of random conformations.

Sets of conformers from the pool whose combined theoretical scattering pattern best fit the experimental data were then selected and compared to the whole pool of conformers. The size distribution of the selected ensemble (Fig. 7a) was narrower than that of the random pool and skewed toward compact conformations. Thus, the results of the EOM analysis are consistent with rigidity of modular connections and FH 10–12 not being significantly flexible or particularly extended. A subsequent EOM iteration was based on the observation of significantly fewer NOEs between CCPs 10 and 11 than between CCPs 11 and 12. Thus, CCPs 11 and 12 along with their linker were defined as a rigid body, while the 10–11 linker was treated as flexible. The resulting narrow size distribution (Fig. 7b) of the selected ensemble suggests that the CCP 10–11 linker is not significantly flexible in the FH 10–12 context, notwithstanding the relatively small surface area buried between modules.

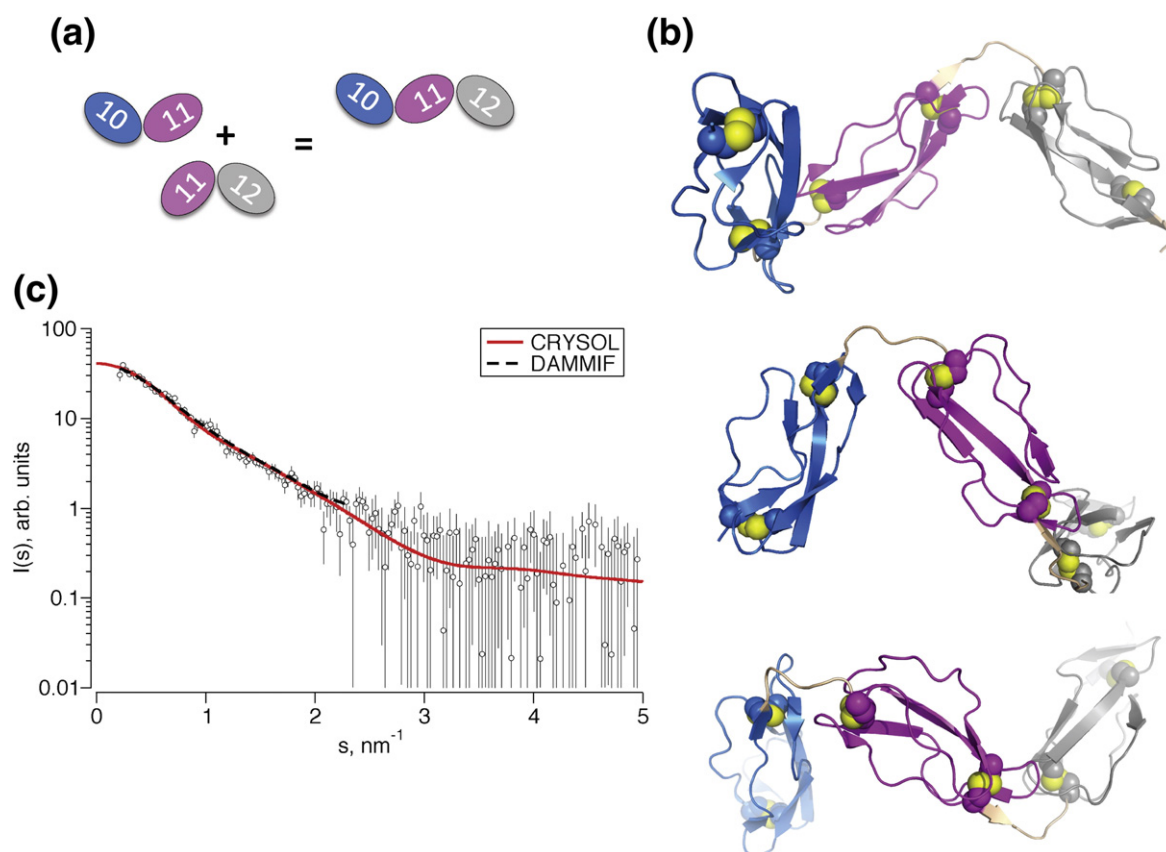


Fig. 6. Structure of FH 10–12. (a) Overview of construction of the FH 10–12 model from concatenation of the FH 10–11 and FH 11–12 NMR structures. (b) Cartoon representation of a model of FH 10–12 built by concatenating structures of FH 10–11 and FH 11–12. Models are rotated about the horizontal axis by 90° and 180° relative to the topmost representation, respectively (for color coding, see Fig. 4). (c) Scattering curve of FH 10–12 fitted to the concatenated structure allows validation of the model.

Structure of FH 10–13 and reinterpreting SAXS data for FH 10–15 and FH 8–15

FH 10–11, FH 11–12 and FH 12–13³⁵ are all relatively rigid, and FH 10–12 lacks significant flexibility. Thus, inter-module angles are likely to be preserved between bi-module and tetra-module contexts. Hence, a concatenated model of FH 10–13—based on the three bi-module structures—should be reliable. MODELLER was used to construct such a model (Fig. 8), which shows how the curvature of FH 10–11 does not follow the arc described by FH 11–13, resulting in an out-of-plane zigzag structure. This model allowed reinterpretation of previously analyzed SAXS data for FH 10–15 and FH 8–15 with a less ambiguous outcome. Given that sequence directionality can be reliably inferred on the basis of the concatenated model of FH 10–13, rigid-body refinement of FH 10–15 consistently yields conformations with a compact C-terminal structure (Fig. 8); in these conformations, CCPs 14 and 15 interact intimately with CCP 13, which could have functional

implications for the regulation of complement by FH as discussed below.

To assess the effect of adding CCPs 8 and 9 to FH 10–15, we conducted EOM analysis on FH 8–15. Lacking information to the contrary, all linkers excluding CCPs 10–11, 11–12 and 12–13 were defined as flexible. It was found that significant populations of both extended and compact inter-module linker conformations were required in order to fit the data, resulting in a broad and bi-modal population distribution (Fig. 7c). Assuming that the FH 10–15 core remains rigid in the context of FH 8–15, this observation implies that CCPs 8 and 9 are not tightly associated with the other six modules. Interestingly, both SAXS data (Fig. 3) and a $[^1\text{H}, ^{15}\text{N}]$ HSQC spectrum collected on recombinant FH 8–9 (data not shown) indicated that this bi-module is as well folded as CCP 11–12 or 12–13. In conclusion, structurally autonomous CCPs 8 and 9 (connected to each other and to CCP 10 with well-conserved linkers) adopt a much less intimate association with CCPs 10–13 than do CCPs 14 and 15; this

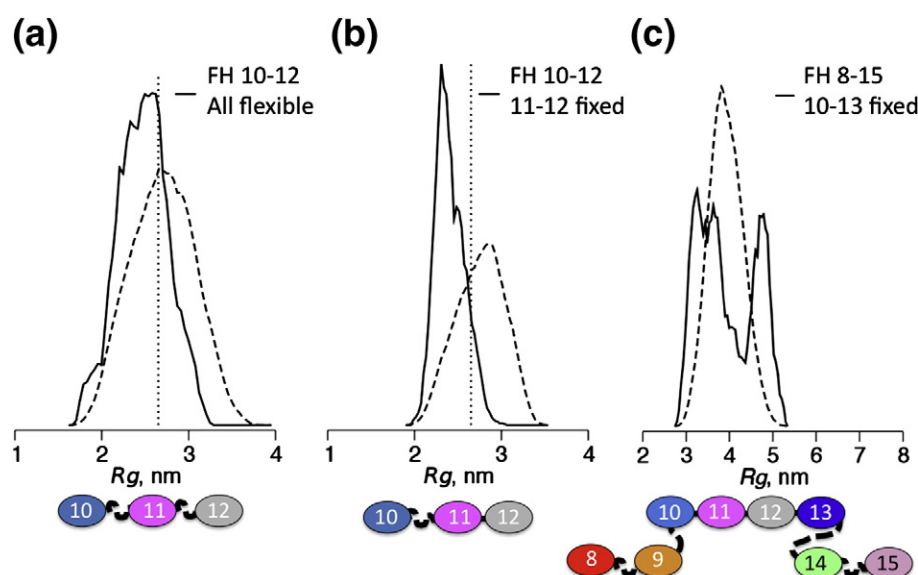


Fig. 7. Applications of the ensemble optimization method. (a) In FH 10–12, individual CCPs were defined as rigid bodies but linkers were allowed to be flexible. Comparison of size distributions shows that an ensemble of structures selected from the pool (continuous line) that fit optimally to the SAXS data for FH 10–12 is more compact than the pool of random structures (broken line). (b) As in (a) except that, in this case, only the CCP 10–11 linker was permitted flexibility. Size distributions are consistent with a rigid arrangement of CCPs 10 and 11 despite limited contacts between modules. (c) An application of EOM to FH 8–15; all individual CCPs along with CCP pairs 10–11, 11–12 and 12–13 were defined as rigid bodies while other linkers were treated as flexible. The selected ensemble indicates a mixture of compact and extended conformers consistent with flexible attachment of CCPs 8 and 9 to a compact, relatively rigid core of CCPs 10–15.

observation supports the idea that the variable 13–14 linker and structural dependency of CCP 14 on its neighbors impose a uniquely compact arrangement of CCPs 14 and 15 with CCPs 10–13.

Discussion

Like all known CCP structures⁶³ (except CCP 13³⁵), CCPs 10, 11 and 12 approximate to flattened prolate spheroids in which N and C termini occupy opposite poles. Given relatively small cross-sectional areas at intersections with neighbors, there are limited ways in which a long string of CCPs (such as FH) could attain the level of rigidity presumably needed for functional cooperation between remote binding sites. The two N-terminal modules of complement receptor type 2, for example, exhibit side-by-side interactions because the eight-residue inter-modular linker is long enough to allow a 142° inter-modular bend (away from linearity).⁵⁴ The rod-like architecture of FH CCP 19–20, on the other hand, arises from its minimal three-residue linker that restricts tilt and ensures interactions between loops on neighboring modules so as to also limit twist.^{30,31}

In the current work on FH 10–11, FH 11–12 and FH 10–12, we describe two CCP–CCP junctions that encompass six-residue inter-modular linkers. By

working in solution, we minimized the risk for observation of potentially anomalous inter-modular tilts such as that observed between CCPs 18 and 19 in the crystal structure of FH 18–20.³⁶ By combining NMR and SAXS, we circumvented the need for measurements of residual dipolar couplings wherein interaction with alignment media carries a theoretical risk of perturbing inter-modular angles or dynamics. The relatively long linkers among CCPs 10, 11 and 12 facilitate large tilts and consequently significant inter-module interactions; their side chains also contribute to the hydrophobic “glue” between adjacent modules. A broadly similar arrangement of modules occurs in FH 12–13.³⁵ Although previous SAXS-based studies of FH 1–5, FH 6–8 and FH 16–20 also indicated bent-back structures,²⁶ high-resolution studies showed FH CCPs 1–3,^{33,34} FH 6–8³² and CCPs 19–20^{36,47,64} to be linear or gently curved. Thus, overlapping bi-modules FH 10–11, FH 11–12 and FH 12–13, connected by linkers of six or eight residues and with tilts of about 80–100°, represent the most strongly tilted succession of four CCPs described to date. The FH conformation is ionic strength dependent and pH sensitive⁶⁵; it is thus worth noting that equivalent His side chains (His651 in FH 11 and His773 in FH 13) feature prominently in both FH 10–11 and FH 12–13 interfaces while a salt bridge, Lys624–Asp675, may stabilize the small FH 10–11 junction.

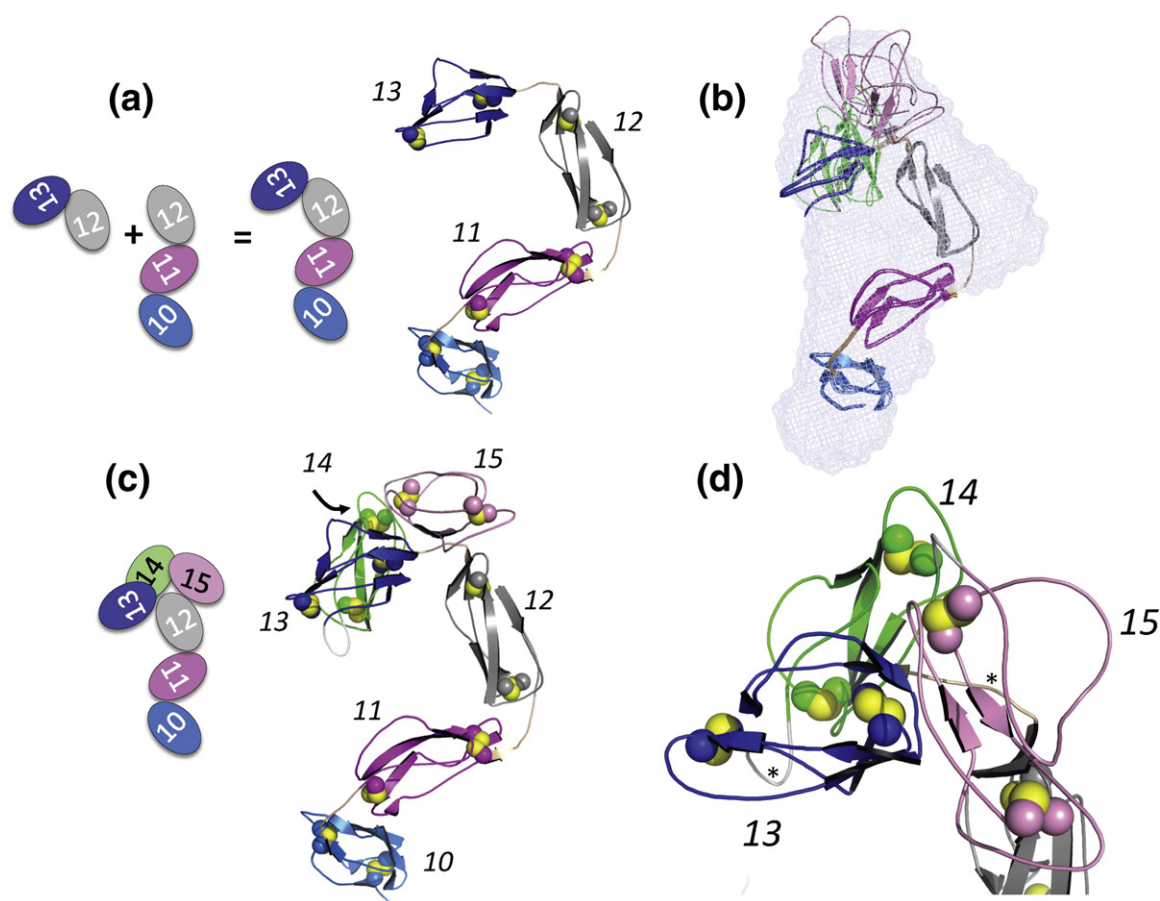


Fig. 8. A model of FH 10–13 helps interpret SAXS data for FH 10–15. (a) Cartoon representation of a concatenated model of FH 10–13. (b) Two typical members of a convergent set of molecular structures that fit well to SAXS data for FH 10–15 interpreted on the assumption that CCPs 10–13 retain the spatial arrangement shown in (a), shown overlaid with the reconstructed SAXS envelope from DAMMIF. (c) A cartoon representation of CCPs 10–15 derived from SAXS data. (d) Zoom-in on CCPs 13–15 of SAXS-derived model; CCPs 14 and 15 bend over the CCP 12–13 linker in this model. Asterisks indicate 13–14 and 14–15 linkers that were built into the fitted structure manually (using MODELLER) since they were treated as dummy atoms during the fitting procedure.

That all three bi-modules exhibit only limited inter-modular flexibility is consistent with the notion that the middle portion of the FH molecule needs flexional reinforcement if the ends (that carry the binding sites) are to be held in (one or more) preferred juxtaposition(s). An assumption that the inter-modular angles observed in the bi-modules 10–11 and 11–12 would persist in the tri-module FH 10–12 was validated by SAXS analysis. Therefore, it was reasonable to concatenate the three bi-modules to reconstruct CCPs 10–13, thus allowing us to trace the path in space of four modules within the central portion of FH. In agreement with a previous analysis of SAXS data for FH 10–15, CCPs 10–13 in this model do not all bend in the same direction to form a horseshoe-like shape. Instead, they form an out-of-plane S-shape or zigzag because the CCP 10–11 bend does not follow the curvature of CCPs 11–13.

We could not extend this instructive NMR-supported exercise to CCPs 14 and 15, as it was impossible to prepare the requisite samples of FH 13–14 or FH 14–15. No suitably stable samples were produced despite repeated attempts to exploit approaches that had previously proved successful for numerous CCP module pairs and longer constructs. This may be a host-specific effect; for example, low stability may arise from the lack of appropriate N-glycosylation⁶⁶ or the presence of a site in folded CCP 14 that is uniquely susceptible to a *Pichia pastoris* protease despite addition of protease inhibitors. Nonetheless, several strands of evidence—described below—suggest that this finding in fact reflects a structural dependency of the common module, that is, CCP 14, on one or more immediate neighbors. Stabilizing interactions between CCP modules were noted in several previous studies,^{67,68} and there are two reports of strand swapping between

neighboring CCP modules^{69,70}; moreover, a non-compactly folded CCP module was discovered in the GABA_B receptor.⁷¹ Thus, instability of CCP 14 in the absence of neighboring modules would not be unprecedented.

In previous SAXS-based studies of FH 10–15, one of two molecular models that fitted the experimental data featured CCP 14 folded back over CCP 13 with extensive, potentially mutually stabilizing, interactions between modules.³⁵ The new solution structures of FH 10–11 and FH 11–12 allow a clearer interpretation of the FH 10–15 SAXS data; importantly, they reinforce the case for a near-180° bend between CCPs 13 and 14. Indeed, in the reinterpreted SAXS and NMR-based model of FH 10–15, CCP 14 contacts not only CCP 13 but also the 12–13 linker (Fig. 8). In the model, CCP 15 also contacts both the 12–13 linker and CCP 13. Notably, recombinant versions of single-module FH 13 and FH 15, but not FH 14, were readily prepared, yielding excellent NMR spectra. Moreover, relative well-behaved samples of FH 11–14, FH 10–15,³⁵ FH 15–18 and FH 15–19⁴⁷ also proved relatively straightforward to prepare. We were additionally able to prepare unlabeled NMR samples of the tri-module FH 13–15 although the quality of the ¹H NMR spectrum was inferior to that of other tri-modules.³⁵

Taking all these observations together implies that, for full stability, CCP 14 requires a stabilizing interface with CCPs 12, 13 and 15. On the other hand, CCPs 13 and 15 do not require CCP 14 (or CCP 12) for stability; rather, based on the lack of appropriate cross-peaks in the HSQC spectrum of FH 13–14, the attachment of putatively disordered CCP 14 to CCP 13, in the module-pair context, actually destabilizes CCP 13. Its highly conserved sequence is not suggestive of CCP 14 being unusual. Indeed, metaPrDOS⁷² does not predict a disordered structure for the CCP 14 sequence. On the other hand, its interface with CCP 13 must indeed be irregular. First, the 13–14 linker is the most variable (between orthologues) of any region within the whole of FH; for example, the CCP 13–14 linker is TSKTS in mice and NEEAKIQL in cows. All other 18 linkers are highly conserved. It is also the only linker of human FH that forms part of an occupied N-glycosylation sequence. Second, CCP 13 is poorly conserved in sequence across orthologues. Third, CCP 13 is less elongated than other CCPs and—importantly—lacks the DE loop and β -strand H that normally help to form a stable end-to-end interface with a subsequent CCP. By contrast, CCPs 8 and 9 that are structurally independent, have (or are predicted to have in the case of CCP 9⁶³) classical CCP structures and are connected to each other and CCP 10 by conserved linkers are flexibly associated with the FH 10–13 unit.

Despite the low resolution of the NMR and SAXS-derived model of CCPs 10–15, it is highly plausible. According to the model, CCPs 12, 13, 14 and 15

form a compact structural unit unlike any other region of FH that has been described to date. CCPs 10, 11, 12 and 13 follow a more open zigzag path with 90° tilts (see Fig. 8). CCPs 12 and 13 act as a bridge between these two structural regions of FH 10–15 by participating in both of them. The linker between CCPs 13 and 14 is very probably unique in the FH molecule in that it has no structural role but forms a flexible and exposed loop on its surface. The positioning of CCP 13 in this model and the fact that CCP 14 is tilted away, leaving the C-terminal end of CCP 13 free, expose not only the non-conserved 13–14 linker but also an extensive highly variable region of CCP 13 encompassing β -strand B and the long BC loop (that includes a helical insertion), the variable DE loop and the aforementioned (non-conserved) N-glycosylation sequon adjacent to the C-terminal cysteine. Note that the presence within the central region of FH of large tilt angles or kinks is consistent, both with best-fit models from SAXS studies²⁵ and with earlier transmission electron micrographs of full-length FH.²⁴

While the sequential series of inter-modular tilts between CCPs 10 and 13 contribute to flexional reinforcement, the purpose of an extreme tilt at 13–14 is less clear. Further studies would be needed to investigate whether this module pair could correspond to the “hinge” in FH that has been suggested³⁵ as a means whereby the molecule can switch rapidly between two conformations—a compact or closed one (as reflected in FH 10–15) that holds the two binding sites in such a way that they have relatively low affinity for C3b⁷³ and an “open” one in which the two ends are able to bind simultaneously, and with high affinity, to nonoverlapping sites on a single molecule of C3b (or adjacent C3b molecules) in the context of a self-surface and its polyanionic markers. More investigation is required to test whether flexional stress or exposure to the electronegative environment found at the host surface disrupts the 13–14 interface. Under these circumstances, module 14 would be likely to become intrinsically disordered, creating a very flexible segment within the FH molecule and allowing terminal regions to cooperate in binding to a target.

Materials and Methods

Protein production

Using procedures similar to those previously described,^{30,74} we produced recombinant proteins in *P. pastoris*. Very briefly, DNA encoding FH 10–11 (amino acid residues 566–687), FH 11–12 (amino acid residues 627–747), FH 10–12 (amino acid residues 566–745), FH 13–14 (amino acid residues 748–865) or FH 14–15 (amino acid residues 808–929) (numbers refer to unprocessed initial human gene product) was PCR amplified from FH

cDNA. Each insert was ligated into a TOPO plasmid vector subsequently used to transform Top10 *Escherichia coli* cells (Invitrogen). The amplified plasmid was digested (PstI and XbaI), and target DNA was ligated into the pPICZα B vector (Invitrogen), 3' of DNA coding for the *Saccharomyces cerevisiae* α-mating factor secretion signal that directs protein production into the secretory pathway. Following amplification in *E. coli* cells and SacI linearization, we used the plasmid to transform *P. pastoris* KM71H. Transformed yeast cells were fermented, and gene expression was induced with methanol. Following centrifugal cell removal, we undertook purification from media using ion-exchange chromatography, typically on SP-Sepharose (GE Healthcare), normally followed by size-exclusion chromatography on a Hi-Load 16/60 Superdex 75 120-mL column (GE Healthcare). Proteins bearing high-mannose N-linked glycans were treated (after first purification step) with Endo H_f (New England Biolabs). Purification was monitored by sodium dodecyl sulfate–polyacrylamide gel electrophoresis under reducing and nonreducing conditions; the molecular weight of the purified material was validated with Fourier transform–ion cyclotron mass spectrometry. For ¹⁵N enrichment, cells were grown in minimal media supplemented with ¹⁵N-labeled ammonium sulfate (Isotec, Sigma-Aldrich), glycerol, basal salts, trace elements and vitamins and were subsequently induced with methanol. For ¹³C, ¹⁵N enrichment, a similar protocol was used but cells were grown in media containing ¹³C glucose (instead of glycerol), and prior to induction with ¹³C methanol, ¹³C glycerol (Isotec, Sigma-Aldrich) was added to facilitate derepression of the alcohol oxidase promoter.

Collection and processing of NMR data and assignment of spectra

NMR spectra were collected at 298 K on samples containing 10% (v/v) D₂O in 5-mm NMR tubes, on Bruker AMX 800- and 600-MHz instruments fitted with cryoprobes. Gradual proteolysis occurred in the case of FH 10–11, necessitating production of three ¹³C, ¹⁵N-labeled NMR samples (at 400, 150 and 500 μM in 20 mM potassium phosphate buffer, pH 6.7). For FH 11–12, one 650-μM ¹³C, ¹⁵N-labeled sample (in 20 mM potassium phosphate buffer, pH 6.3) sufficed. In the case of FH 13–14, production of a properly folded and stable sample suitable for NMR proved difficult (see Results) and only a ¹⁵N-labeled sample (150 μM in 20 mM potassium phosphate buffer, pH 6.6) was prepared. Topspin version 1.3 (Bruker) was used to set up the experiments, to acquire data and for initial data processing. The data were brought to their final format by processing using AZARA (W. Boucher, Department of Biochemistry, University of Cambridge, UK).

Resonance assignments were accomplished using the CCPN Analysis software.⁷⁵ For backbone assignments, HBHANH (for FH 11–12), HBHA(CO)NH, CBCANH, CBCA(CO)NH, HNCO and HN(CA)CO experiments were collected. For side-chain assignments, HCCH–total correlated spectroscopy (TOCSY), H(C)(CO)NH–TOCSY and (H)C(CO)NH–TOCSY experiments were acquired along with (two-dimensional) (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments for assigning aromatic side chains.

Solution structure determination

Including the nonnative N-terminal sequence (EAAG and EAEAG in FH 10–11 and FH 11–12, respectively), 88% of triple-resonance assignments, including 97% of backbone atoms for FH 10–11 and 88% of triple-resonance assignments, including 95% of backbone atoms for FH 11–12, were completed. Of the theoretically assignable backbone atoms, the following were missing: Pro632 (CO), Pro633 (CO) and Asp675 (N and HN) in the case of FH 10–11 and Glu621 (H, HN, CO, C^α), Ala622 (H, HN and C^α), Pro632 (CO, C^α), Pro633 (CO), Asp675 (N, HN), Pro707 (CO) and Ile747 (CO) in the case of FH 11–12.

In one out of twelve in the case of FH 10–11 (Pro618) and one out of ten in the case of FH 11–12 (Pro708), X-Pro linkages were defined as *cis*, while the remainder were defined as *trans* as judged by chemical shift differences between Pro ¹³C^β and ¹³C^γ atoms,⁷⁶ by strong NOEs observed between H^α (Xaaⁱ⁻¹) and H^α (Proⁱ) and between H^α (Xaaⁱ⁻¹) and HN (Xaaⁱ⁺¹) and by no or weak detectable NOEs between H^α (Xaaⁱ⁻¹) and H^δ (Proⁱ). Based on mass spectrometry, all eight cysteine residues in both FH 10–11 and FH 11–12 were inferred to be in the oxidized state. Specific disulfide bridges were incorporated into the final rounds of structure calculations on the basis of NOEs (reinforced by a precedent established by previous studies of CCPs). Approximately 3% (for FH 10–11) and 14% (for FH 11–12) of peaks in the ¹⁵N-edited NOE spectroscopy spectra were assigned manually while, for both constructs, ~8% of peaks in the ¹³C-edited NOE spectroscopy spectra were also assigned manually.

Remaining NOEs were assigned by a combination of automated assignment and structure calculation. Thus, seven cycles of CYANA 2.1⁴⁹ were employed for each of FH 10–11 and FH 11–12 to generate lists of NOE-derived distance restraints and sets of preliminary structures. Once CYANA-generated parameters met published quality criteria,⁴⁹ distance restraints were transferred into the program CNS,⁵⁰ using the CCPN Format Converter program. Working within CNS allows structure refinement against explicit water.

Assessing and comparing structures

The quality of structures was assessed using PROCHECK⁷⁷ (see Table 1). A hydrogen–deuterium exchange experiment was performed to cross-validate hydrogen bonds observed in the NOE-derived structures; for this purpose, lyophilized protein was transferred to deuterated buffer, whereupon slowly exchanging amide protons were identified in a [¹⁵N, ¹H]HSQC spectrum collected 15 min after exposure to D₂O. Secondary structure elements were identified using STRIDE.⁵¹

The Web server VADAR 1.8⁷⁸ was used to establish the solvent accessible surface area (Å²); the buried surface area was computed as (SA module_i + SA module_j) – SA bi-module_{ij}, where CCP *i* was considered to encompass one residue before its Cys(I) and three residues after its Cys(IV), and CCP (*j*) boundaries were considered the third residue before its Cys(I) and one residue after its Cys(IV).

Inter-modular tilt, twist and skew angles^{29,79} were determined using an in-house script[†]; a vector was chosen between the principal axis of the inertia tensor (the z-axis)

and the C $^{\alpha}$ of Leu616 (CCP 10), Trp678 (CCP 11) or Trp738 (CCP 12), respectively, and with module boundaries defined as Cys(I), that is, Cys569 (CCP 10), Cys630 (CCP 11) or Cys691 (CCP 12), and Cys(IV), that is, Cys623 (CCP 10), Cys684 (CCP 11) or Cys744 (CCP 12), respectively.

Combinatorial extension⁸⁰ was employed to compare each experimentally determined CCP structure within the complement system against the closest-to-mean individual structures determined herein.

SAXS data collection and interpretation

Synchrotron radiation X-ray scattering data were collected on the X33 beamline of the European Molecular Biology Laboratory (Deutsches Elektronen-Synchrotron, Hamburg) using a 1M PILATUS pixel detector (Dectris, Switzerland) and eight frames of 15-s exposure time. Samples were analyzed at 20 °C, using protein concentrations of 0.3–10.7 mg/mL in phosphate-buffered saline. The sample-to-detector distance was 2.7 m, covering a range of momentum transfer $0.08 \leq s \leq 6.0 \text{ nm}^{-1}$ [where $s = 4\pi \sin(\theta)/\lambda$, with 2θ being the scattering angle and $\lambda = 0.15 \text{ nm}$ being the X-ray wavelength]. Based on comparison of successive 15-s frames, samples in which radiation damage was detected were further scrutinized and only frames showing no significant change in intensity were used for subsequent data analysis.

Data from the detectors were normalized to the transmitted beam intensity and averaged, and the scattering of buffer solutions was subtracted. The difference curves were scaled for solute concentration. Data manipulations were performed using the ATSAS software package^{81,82} utilizing the recently developed DATTOOL library for averaging, merging and extrapolation to infinite dilution from concentration series (Guinier regions for samples at several concentrations, showing linearity; Supplementary Fig. S8). The forward scattering $I(0)$ and radius of gyration R_g were determined from Guinier analysis,⁸³ assuming that, at very small angles ($s \leq 1.3/R_g$), the intensity is represented as $I(s) = I(0)\exp(-(sR_g)^2/3)$. These parameters were also estimated from the full scattering curves using the indirect Fourier transform method implemented in the program GNOM,⁸⁴ along with the distance distribution function $p(r)$ and the maximum particle dimensions D_{max} . Molecular masses of solutes were estimated from SAXS data by comparing the extrapolated forward scattering with that of a reference solution of bovine serum albumin (66 kDa).

Fit of the NMR structures of FH fragments to the SAXS data

The fits of the NMR-derived structures of FH fragments to the SAXS data were conducted using the program CRY SOL.⁵⁷ CRY SOL calculates the partial scattering amplitudes of proteins from their atomic coordinates, taking into account the hydration layer and excluded solvent volume. The number of harmonics was set to 50 and 256 points computed across the entire data range ($0 < s < 5 \text{ nm}^{-1}$). The NMR-derived structures were also fitted to the SAXS data with Xplor-NIH using 500 solid angles and 100 data points evenly distributed in a cubic spline constructed using the SAXS data within the range specified above.⁸⁵

Analysis of flexibility

Analysis of inter-module flexibility and size distribution of possible conformers consistent with the measured scattering data was conducted using the EOM.⁶² The EOM selects an ensemble of possible conformations of input rigid bodies (using the CCP modules 10, 11 and 12 determined in this study and the remaining modules as described previously³⁴) and user-defined regions of flexible sequence (here defined as the linkers between CCP modules) from a pool of 10,000 randomly generated models, using CRY SOL to calculate the theoretical scattering profiles and a genetic algorithm, GAJOE, to select the representative set that best describes the experimental data.

Ab initio shape determination and molecular modeling

Low-resolution-shape envelopes were determined using the *ab initio* bead-modeling program DAMMIF.⁵⁸ DAMMIF represents the particle as a collection of M densely packed beads inside an adaptable and loosely constrained search volume compatible with the experimentally determined R_g . Each bead is randomly assigned to solvent (index=0) or solute (index=1), and the particle structure in solution is described by a binary string of length M . Disconnected strings of beads are rejected, and the scattering amplitudes are calculated. Simulated annealing is then used to search for a compact model that minimizes the discrepancy between the experimental and calculated intensities of momentum transfer. The results of 10 independent DAMMIF reconstructions were compared using SUPCOMB13⁵⁹ to determine the most representative (typical) model. Averaged DAMMIF models were also determined using DAMAVER, and these models were adjusted such that they agree with the experimentally determined excluded volume using DAMFILT.⁸⁶

Molecular modeling used, as rigid bodies and where appropriate, the same structures as used for EOM. Rigid-body models were generated using the program CORAL, an advanced version of the rigid-body modeling program BUNCH^{81,87} where linkers/loops between individual subunits are represented as random polypeptide chains. The results of 10 independent CORAL runs were analyzed using the programs SUPCOMB13 and DAMAVER to identify the most representative/typical models.

Combining SAXS and NOEs

A concatenated FH 10–12 NOE list was created by merging experimentally derived FH 10–11 and FH 11–12 NOE lists. Only NOEs from the FH 10–11 list that involve residues in (and between) CCP 10, the CCP 10–11 linker and the N-terminal half of the three-dimensional structure of CCP 11 were used; likewise, the only NOEs extracted for use from the FH 11–12 list were those in (and between) the C-terminal half of the three-dimensional structure of CCP 11, the CCP 11–12 linker and CCP 12. This procedure, similar to that described previously,⁵⁶ was designed to remove potential duplications and conflict involving the common CCP 11.

Structure calculations, incorporating the SAXS data for FH 10–12 and the concatenated NOE list, were carried out

in Xplor-NIH version 2.30.⁵⁵ A naive model of FH 10–12, prepared using MODELLER 9v10⁵⁸ by overlaying closest-to-mean structures of FH 10–11 and FH 11–12 on the common CCP 11, was used as a starting structure. In the first stage of a simulated annealing protocol,⁵⁵ the temperature was lowered from an initial 2000 K to 600 K in 50-K intervals, with 300 steps of molecular dynamics at each temperature increment; within this temperature interval, the energy terms were progressively introduced by multiplicative ramping. In the second stage, the structures were cooled from 600 K to 100 K, in increments of 25 K, with 300 steps of molecular dynamics at each temperature; potential terms were statically set to the top values used in the first stage of cooling. Two calculations were carried out: first, a set of 100 standard protein structures was calculated; second, 100 ensembles of structures (with number of members $N=2$ in each ensemble) were calculated. In the ensemble calculation, the energy terms were calculated as an ensemble average, as described before.⁶¹ Similar calculations were performed for FH 10–11 and FH 11–12, using the relevant SAXS data and NOE lists and NOE-only derived structures as starting structures.

Relaxation data

Measurements of backbone ^{15}N T_1 and T_2 and ^1H , ^{15}N NOEs were conducted using a 600-MHz magnet and 0.45-nM and 0.65-mM (for FH 10–11 and FH 11–12, respectively) samples in 20 mM potassium phosphate buffer, pH 6.7 (FH 10–11) or pH 6.3 (FH 11–12). Delays used for T_1 (ms) were (FH 10–11) 51.2, 51.2, 301.2, 501.2, 601.2, 751.2, 851.2 and 901.2 or (FH 11–12) 51.2, 301.2, 501.2, 601.2, 701.2, 801.2 and 901.2. Delays used for T_2 (ms) (FH 10–11) were 16.96, 16.96, 33.92, 67.84, 101.76, 118.72, 135.68 and 152.64 or (FH 11–12) 16.96, 33.92, 67.84, 84.8, 101.76, 118.82 and 135.68. For heteronuclear NOE measurements, a reference experiment with a 5-s relaxation delay was followed by a second spectrum recorded with ^1H saturation achieved by a train of 120° pulses applied for the last 3 s of the 5-s delay to attain NOE build up. The NMR data were processed using AZARA, and spectra were assigned using Analysis.⁷⁵ Relaxation rates were obtained by fitting a single-exponential decay to the extracted cross-peak height of each residue using nonlinear fitting. In the case of FH 10–11, the following native residues were excluded from the analyses: Glu566, Arg567, Lys583, His613, Ser652 and Gly676 (signals from the backbone amides were too weak); Asp581, Val589, Val592, Phe595, Cys597, Phe601, Val609, Val627, Thr645, Tyr649 and Ile671 (due to overlap); and Asp675 (amide not found). In the case of FH 11–12, the following native residues were excluded: Ser652, His735 and Gly736 (weak signals); Tyr658, Leu697, Gln703, Tyr709, Ile731 and Cys733 (due to overlap); and Asp675 (amide not found).

Accession numbers

NMR data and structure coordinates have been deposited at the BMRB (Biological Magnetic Resonance Bank)

and at the PDB (Protein Data Bank). Accession numbers are as follows:

FH 10–11: PDB, 4b2r; BMRB, 18604

FH 11–12: PDB, 4b2s; BMRB, 18599

Acknowledgements

This work was funded by the Wellcome Trust grants 073131 and 085025. H.D.T.M. was supported by a fellowship from the European Molecular Biology Laboratory Interdisciplinary Postdoctoral Programme. Use of the Edinburgh Protein Production Facility was supported by The Wellcome Trust, the Scottish University Life Sciences Alliance and the Biotechnology and Biological Sciences Research Council.

Supplementary Data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2012.09.013>

Received 7 August 2012;

Received in revised form 17 September 2012;

Accepted 18 September 2012

Available online 25 September 2012

Keywords:

protein NMR;
protein domains;
complement system;
small-angle X-ray scattering;
regulators of complement activation

† Available from www.bionmr.chem.ed.ac.uk/bionmr/public_html/ccp-db.html

Present addresses: H. D. T. Mertens, Australian Synchrotron, 800 Blackburn Road, Clayton VIC 3168, Australia; C. Q. Schmidt, Institute of Pharmacology of Natural Products and Clinical Pharmacology, Ulm University, Helmholtzstrasse 20, D-89081 Ulm, Germany.

Abbreviations used:

CCP, complement control protein; CR1, complement receptor type 1; DAF, decay accelerating factor; FH, factor H; EOM, ensemble optimization method; HSQC, heteronuclear single quantum coherence; MCP, membrane cofactor protein; NOE, nuclear Overhauser enhancement; SAXS, small-angle X-ray scattering; TOCSY, total correlated spectroscopy.

References

- Walport, M. J. (2001). Complement. First of two parts. *N. Engl. J. Med.* **344**, 1058–1066.
- Walport, M. J. (2001). Complement. Second of two parts. *N. Engl. J. Med.* **344**, 1140–1144.
- Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. (2010). Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* **11**, 785–797.
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. (1981). Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med.* **154**, 856–867.
- Lachmann, P. J. (2009). The amplification loop of the complement pathways. *Adv. Immunol.* **104**, 115–149.
- Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L. & Pahl, J. W. (1980). Evidence for presence of an internal thiolester bond in third component of human complement. *Proc. Natl Acad. Sci. USA*, **77**, 5764–5768.
- Kirkitadze, M. D. & Barlow, P. N. (2001). Structure and flexibility of the multiple domain proteins that regulate complement activation. *Immunol. Rev.* **180**, 146–161.
- Meri, S. & Pangburn, M. K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl Acad. Sci. USA*, **87**, 3982–3986.
- Rosen, H. & Law, S. K. (1990). The leukocyte cell surface receptor(s) for the iC3b product of complement. *Curr. Top. Microbiol. Immunol.* **153**, 99–122.
- Esparza-Gordillo, J., Soria, J. M., Buil, A., Almasy, L., Blangero, J., Fontcuberta, J. & Rodriguez de Cordoba, S. (2004). Genetic and environmental factors influencing the human factor H plasma levels. *Immunogenetics*, **56**, 77–82.
- Weiler, J. M., Daha, M. R., Austen, K. F. & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein β 1H. *Proc. Natl Acad. Sci. USA*, **73**, 3268–3272.
- Sim, R. B. & DiScipio, R. G. (1982). Purification and structural studies on the complement-system control protein β 1H (factor H). *Biochem. J.* **205**, 285–293.
- de Cordoba, S. R. & de Jorge, E. G. (2008). Translational mini-review series on complement factor H: genetics and disease associations of human complement factor H. *Clin. Exp. Immunol.* **151**, 1–13.
- Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I. *et al.* (2005). A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA*, **102**, 7227–7232.
- Pangburn, M. K. (2000). Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology*, **49**, 149–157.
- Schmidt, C. Q., Herbert, A. P., Hocking, H. G., Uhrin, D. & Barlow, P. N. (2008). Translational mini-review series on complement factor H: structural and functional correlations for factor H. *Clin. Exp. Immunol.* **151**, 14–24.
- Brook, E., Herbert, A. P., Jenkins, H. T., Soares, D. C. & Barlow, P. N. (2005). Opportunities for new therapies based on the natural regulators of complement activation. *Ann. N. Y. Acad. Sci.* **1056**, 176–188.
- Smith, R. J., Alexander, J., Barlow, P. N., Botto, M., Cassavant, T. L., Cook, H. T. *et al.* (2007). New approaches to the treatment of dense deposit disease. *J. Am. Soc. Nephrol.* **18**, 2447–2456.
- Ricklin, D. & Lambris, J. D. (2007). Complement-targeted therapeutics. *Nat. Biotechnol.* **25**, 1265–1275.
- Reid, K. B. & Day, A. J. (1989). Structure–function relationships of the complement components. *Immunol. Today*, **10**, 177–180.
- Ripoche, J., Day, A. J., Willis, A. C., Belt, K. T., Campbell, R. D. & Sim, R. B. (1986). Partial characterization of human complement factor H by protein and cDNA sequencing: homology with other complement and non-complement proteins. *Biosci. Rep.* **6**, 65–72.
- Ripoche, J., Day, A. J., Harris, T. J. & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochem. J.* **249**, 593–602.
- Perkins, S. J., Nealis, A. S. & Sim, R. B. (1991). Oligomeric domain structure of human complement factor H by X-ray and neutron solution scattering. *Biochemistry*, **30**, 2847–2857.
- DiScipio, R. G. (1992). Ultrastructures and interactions of complement factors H and I. *J. Immunol.* **149**, 2592–2599.
- Aslam, M. & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *J. Mol. Biol.* **309**, 1117–1138.
- Okemefuna, A. I., Gilbert, H. E., Griggs, K. M., Ormsby, R. J., Gordon, D. L. & Perkins, S. J. (2008). The regulatory SCR-1/5 and cell surface-binding SCR-16/20 fragments of factor H reveal partially folded-back solution structures and different self-associative properties. *J. Mol. Biol.* **375**, 80–101.
- Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B. & Campbell, I. D. (1991). Secondary structure of a complement control protein module by two-dimensional ^1H NMR. *Biochemistry*, **30**, 997–1004.
- Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B. & Campbell, I. D. (1991). Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* **219**, 717–725.
- Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B. & Campbell, I. D. (1993). Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.* **232**, 268–284.
- Herbert, A. P., Uhrin, D., Lyon, M., Pangburn, M. K. & Barlow, P. N. (2006). Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in 3D structure. *J. Biol. Chem.* **281**, 16512–16520.
- Jokiranta, T. S., Jaakola, V. P., Lehtinen, M. J., Parepalo, M., Meri, S. & Goldman, A. (2006). Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. *EMBO J.* **25**, 1784–1794.
- Prosser, B. E., Johnson, S., Roversi, P., Herbert, A. P., Blaum, B. S., Tyrrell, J. *et al.* (2007). Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* **204**, 2277–2283.

33. Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K. *et al.* (2008). Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J. Biol. Chem.* **283**, 9475–9487.
34. Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J., Lambris, J. D. & Gros, P. (2009). Structure of complement fragment C3b–factor H and implications for host protection by complement regulators. *Nat. Immunol.* **10**, 728–733.
35. Schmidt, C. Q., Herbert, A. P., Mertens, H. D., Guariento, M., Soares, D. C., Uhrin, D. *et al.* (2010). The central portion of factor H (modules 10–15) is compact and contains a structurally deviant CCP module. *J. Mol. Biol.* **395**, 105–122.
36. Morgan, H. P., Mertens, H. D., Guariento, M., Schmidt, C. Q., Soares, D. C., Svergun, D. I. *et al.* (2012). Structural analysis of the C-terminal region (modules 18–20) of complement regulator factor H (FH). *PLoS One*, **7**, e32187.
37. Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S. *et al.* (2008). A new map of glycosaminoglycan and C3b binding sites on factor H. *J. Immunol.* **181**, 2610–2619.
38. Pechtl, I. C., Neely, R. K., Dryden, D. T., Jones, A. C. & Barlow, P. N. (2011). Use of time-resolved FRET to validate crystal structure of complement regulatory complex between C3b and factor H (N terminus). *Protein Sci.* **20**, 2102–2112.
39. Alsenz, J., Lambris, J. D., Schulz, T. F. & Dierich, M. P. (1984). Localization of the complement-component-C3b-binding site and the cofactor activity for factor I in the 38 kDa tryptic fragment of factor H. *Biochem. J.* **224**, 389–398.
40. Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J. & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *J. Immunol.* **155**, 348–356.
41. Sharma, A. K. & Pangburn, M. K. (1996). Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc. Natl Acad. Sci. USA*, **93**, 10996–11001.
42. Kuhn, S. & Zipfel, P. F. (1996). Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur. J. Immunol.* **26**, 2383–2387.
43. Pangburn, M. K., Atkinson, M. A. & Meri, S. (1991). Localization of the heparin-binding site on complement factor H. *J. Biol. Chem.* **266**, 16847–16853.
44. Blackmore, T. K., Hellwage, J., Sadlon, T. A., Higgs, N., Zipfel, P. F., Ward, H. M. & Gordon, D. L. (1998). Identification of the second heparin-binding domain in human complement factor H. *J. Immunol.* **160**, 3342–3348.
45. Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N. & Pangburn, M. K. (2006). Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J. Immunol.* **177**, 6308–6316.
46. Ferreira, V. P., Herbert, A. P., Cortes, C., McKee, K. A., Blaum, B. S., Esswein, S. T. *et al.* (2009). The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome. *J. Immunol.* **182**, 7009–7018.
47. Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P. *et al.* (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol.* **18**, 463–470.
48. Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riande, S., Siret, L. & Bihoreau, N. (2007). Site-specific N-glycan characterization of human complement factor H. *Glycobiology*, **17**, 932–944.
49. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378.
50. Brunger, A. T. (2007). Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733.
51. Heinig, M. & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502.
52. Smith, B. O., Mallin, R. L., Krych-Goldberg, M., Wang, X., Hauhart, R. E., Bromek, K. *et al.* (2002). Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor. *Cell*, **108**, 769–780.
53. Persson, B. D., Schmitz, N. B., Santiago, C., Zocher, G., Larvie, M., Scheu, U. *et al.* (2010). Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. *PLoS Pathog.* **6**, e1001122.
54. Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M. & Chen, X. S. (2001). Structure of complement receptor 2 in complex with its C3d ligand. *Science*, **292**, 1725–1728.
55. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. (2006). Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62.
56. Maciejewski, M., Tjandra, N. & Barlow, P. N. (2011). Estimation of interdomain flexibility of N-terminus of factor H using residual dipolar couplings. *Biochemistry*, **50**, 8138–8149.
57. Svergun, D., Barberato, C. & Koch, M. H. J. (1995). CRYSOLE—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
58. Franke, D. & Svergun, D. I. (2009). DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346.
59. Kozin, M. B. & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.
60. Schwieters, C. D., Suh, J. Y., Grishaev, A., Ghirlando, R., Takayama, Y. & Clore, G. M. (2010). Solution structure of the 128 kDa enzyme I dimer from *Escherichia coli* and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. *J. Am. Chem. Soc.* **132**, 13026–13045.
61. Clore, G. M. & Schwieters, C. D. (2004). How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J. Am. Chem. Soc.* **126**, 2923–2938.
62. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664.
63. Soares, D. C., Gerloff, D. L., Syme, N. R., Coulson, A. F., Parkinson, J. & Barlow, P. N. (2005). Large-scale modelling as a route to multiple surface comparisons of the CCP module family. *Protein Eng., Des. Sel.* **18**, 379–388.

64. Kajander, T., Lehtinen, M. J., Hyvarinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E. *et al.* (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host–nonhost discrimination by complement. *Proc. Natl Acad. Sci. USA*, **108**, 2897–2902.
65. Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. (2009). Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol.* **391**, 98–118.
66. Fenaille, F., Le Mignon, M., Groseil, C., Siret, L. & Bihoreau, N. (2007). Combined use of 2,4,6-trihydroxyacetophenone as matrix and enzymatic deglycosylation in organic-aqueous solvent systems for the simultaneous characterization of complex glycoproteins and N-glycans by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**, 812–816.
67. Kirkitadze, M. D., Dryden, D. T., Kelly, S. M., Price, N. C., Wang, X., Krych, M. *et al.* (1999). Co-operativity between modules within a C3b-binding site of complement receptor type 1. *FEBS Lett.* **459**, 133–138.
68. Kirkitadze, M. D., Krych, M., Uhrin, D., Dryden, D. T., Smith, B. O., Cooper, A. *et al.* (1999). Independently melting modules and highly structured intermodular junctions within complement receptor type 1. *Biochemistry*, **38**, 7019–7031.
69. Rickert, M., Wang, X., Boulanger, M. J., Goriatcheva, N. & Garcia, K. C. (2005). The structure of interleukin-2 complexed with its alpha receptor. *Science*, **308**, 1477–1480.
70. Andersen, C. B., Torvund-Jensen, M., Nielsen, M. J., de Oliveira, C. L., Hersleth, H. P., Andersen, N. H. *et al.* (2012). Structure of the haptoglobin–haemoglobin complex. *Nature*, **489**, 456–459.
71. Blein, S., Ginham, R., Uhrin, D., Smith, B. O., Soares, D. C., Veltel, S. *et al.* (2004). Structural analysis of the complement control protein (CCP) modules of GABA(B) receptor 1a: only one of the two CCP modules is compactly folded. *J. Biol. Chem.* **279**, 48292–48306.
72. Ishida, T. & Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, **24**, 1344–1348.
73. Oppermann, M., Manuelli, T., Jozsi, M., Brandt, E., Jokiranta, T. S., Heinen, S. *et al.* (2006). The C-terminus of complement regulator Factor H mediates target recognition: evidence for a compact conformation of the native protein. *Clin. Exp. Immunol.* **144**, 342–352.
74. Schmidt, C. Q., Slingsby, F. C., Richards, A. & Barlow, P. N. (2011). Production of biologically active complement factor H in therapeutically useful quantities. *Protein Expression Purif.* **76**, 254–263.
75. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M. *et al.* (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
76. Schubert, M., Labudde, D., Oschkinat, H. & Schmieider, P. (2002). A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics. *J. Biomol. NMR*, **24**, 149–154.
77. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
78. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **31**, 3316–3319.
79. Soares, D. C. & Barlow, P. N. (2005). Complement control protein modules in the regulators of complement activation. In *Structural Biology of the Complement System* (Morikis, D. & Lambris, J. D., eds), pp. 19–62, CRC Press, Taylor & Francis Group, Boca Raton, FL.
80. Shindyalov, I. N. & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747.
81. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M. *et al.* (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350.
82. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
83. Guinier, A. (1939). La diffraction des rayons X aux tres petits angles: application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris)*, **12**, 161–237.
84. Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503.
85. Schwieters, C. D. & Clore, G. M. (2007). A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry*, **46**, 1152–1166.
86. Volkov, V. V. & Svergun, D. I. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.
87. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250.
88. Eswar, N., Eramian, D., Webb, B., Shen, M. Y. & Sali, A. (2008). Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426**, 145–159.

Glossary

Complement: Approximately 30 proteins involved in multiple aspects of innate and acquired immunity.

Alternative pathway: Complement activation via spontaneous C3 hydrolysis; in contrast, antibody–antigen complexes and polysaccharides are needed to activate the classical and lectin pathways, respectively.

C3: A complement protein that is cleaved by C3 convertase into C3a and C3b; cleavage activates a thioester group in C3b that is able to react with any nearby nucleophile.

C3(H₂O): A form of C3 in which the thioester has been spontaneously hydrolyzed; resembles C3b in domain arrangements and ability to form part of a C3 convertase complex.

C3 convertase: Binary complex—C3(H₂O).Bb or C3b.Bb—that cleaves C3 to C3a and C3b; formation involves C3b or C3(H₂O) binding proenzyme factor B that subsequently gets cleaved (and activated) by factor D to Bb.

CCP modules: Domains, which predominate among regulators of complement activation, containing ~60 amino acid residues and two disulfides.